Terminal Digits and the Examination of Questioned Data

James E. Mosimann, Office of Research Integrity, DHHS, USA

John E. Dahlberg, Office of Research Integrity, DHHS, USA

Nancy M. Davidian, Office of Research Integrity, DHHS, USA

John W. Krueger, Office of Research Integrity, DHHS, USA

Keywords. Digits, Research misconduct, Statistical forensics, Terminal Digits, Uniform distribution

Our objective is to illustrate the use of statistical methods to examine the authenticity of data in the investigation of research misconduct. We present examples of statistical analyses of questioned data from several cases that illustrate the experience of the Office of Research Integrity. We show that the statistical examination of numbers that are normally *unrepeatable* when experiments are repeated, or otherwise are of inconsequential meaning, may reveal substantial clues as to the authenticity of questioned data when compared with numbers in data that are unquestioned. We illustrate the occurrence of the uniform distribution of non-leading (insignificant rightmost) digits in unquestioned numbers, along with examples of deviation from such uniformity for fabricated or falsified numbers. (Most people are unable to choose digits randomly.) We describe several cases in which a variety of anomalies in data sets provided the impetus for the examination of rightmost digits. The anomalous behavior of rightmost digits, when added to testimony and other physical evidence, can greatly enhance or decrease the credibility of witnesses. The cases discussed involve: 1 and 2, Anomalous behavior of terminal digits in published or recorded numbers; 3, Terminal *odd* digits in event times that should have exhibited only *even* digits (and why); and 4, Data that were falsified by calculations from computer spreadsheets (detected by the inclusion of an additional digit of accuracy).

Introduction

Allegations of research misconduct[1] often are of the form that a particular experiment was not done as described, or not done at all. In considering such allegations it is often necessary to examine "questioned" data. Such data can establish that the experiment was performed as described. However, if the allegation is true, then these questioned data are necessarily falsified or fabricated.

A useful way to assess questioned data is to examine inconsequential components of data sets that are not directly related to the scientific conclusions of the purported experiment. Thus if the allegation is true and the data are falsified, the falsifier typically devotes attention to numbers that establish the desired scientific outcome. Properties of the numbers that are not directly related to the desired outcome are less likely to receive consideration by the falsifier.

The same principle of examining details inconsequential to the scientific outcome appears valid

Corresponding author: J.E. Mosimann (present address), ABL Associates Inc., 5 Balmoral Court, Rockville, MD 20850, jemosimann@erols.com.

whether the data are expressed in non-numeric form (images, written descriptions) or as numbers. Here we consider several cases where the data are numeric and lend themselves to immediate statistical description.

In all these cases we stress the importance of comparing "questioned" data with similar unquestioned data from the same laboratory or individuals.

Rightmost digits

Consider counts of radioactivity for a biological preparation; for example, 5071. In a recount of the sample, or in a replication of the assay, it is highly unlikely that the rightmost digits will be the same. Thus with two repetitions of the experimental procedure, instead of 5071, one might obtain respectively, 5109 and 4966. The rightmost, non-leading digits of these three numbers are not the same. Thus _071 differs from _109, and in turn both differ from _966.

Digits are often recorded well beyond the repeatability of the experimental procedure. For such rightmost digits, theoretically[2] there is a tendency to be uniformly distributed as expected in a lottery. For example, a uniform distribution of digits is expected in the Maryland Lottery. Figure 1 shows the frequencies of the digits 0 to 9 found in 5,106 winning "Pick-3" numbers (of 3 digits each) for the past ten years.[3] This distribution is not significantly different from uniform. All digits have occurred with nearly the same frequency, as they should in a lottery.

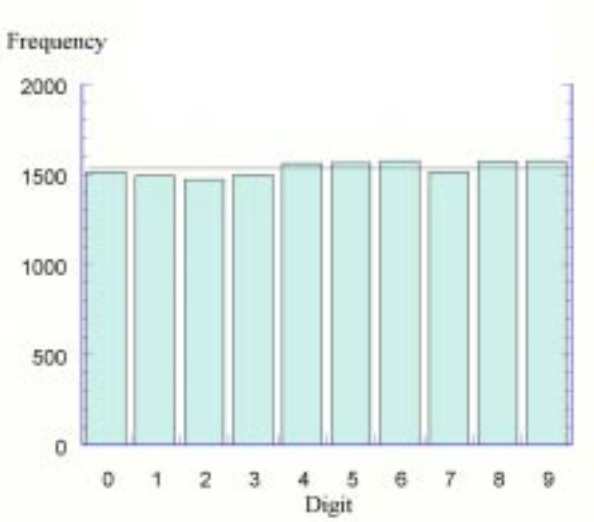

Figure 1. Ten years of Maryland Lottery Pick Three Digits, January 2, 1990 to December 31; 15,318 digits.

Case 1: Uniformly distributed rightmost digits in scintillation counts

In the first case, experimental measurements were known not to have been done because radioactive spots on the experimental sheets had not been excised and hence could not have been counted in the scintillation counter. Yet the respondent's notebook contained (falsified) handwritten counts for that experiment. In this case, faced with the evidence, the respondent admitted to the falsification of the numbers in the notebook.

In addition to the questioned counts, the notebook contained handwritten counts that were supported by counter output, and thus not falsified. Both questioned and unquestioned numbers occur in pairs (a numerator and denominator) and have large numbers of digits (Table 1).

The following procedure was used to find digits. The rightmost digit of a number was designated as occupying "Place 1," then the digit to its left occupied "Place 2," etc. Digits were examined in four places for each number, except that the leftmost digit was never included in the analysis. Thus by way of example, the underlined digits would be included in the analysis: 1<u>078</u>, 25<u>1183</u>, 7<u>35</u>, 6<u>2034</u>. It is clear that a three-digit number contributes two digits for analysis and a four-digit number, three digits. Numbers of five or more digits contribute four digits.

Chi-Square tests for uniformity of digit distributions from 252 falsified counts from notebook pages 141-152 are presented in Table 2. The distributions are not uniform. Three of the four Chi-Square values have probabilities less than .05, and when digits from all four places are grouped together, the total distribution is far from uniform (Chi-Square = 30.94, df = 9, p=.0003).

Chi-Square tests for uniformity of the digit distributions from 222 unquestioned counts also are presented in Table 2. The distributions are not significantly different from uniform. All of the four Chi-Square values have probabilities greater than .05, and when digits from all four places are grouped together, the total distribution is not significantly different from uniform (Chi-Square = 11.09, df = 9, p=.27).

The unquestioned counts have uniform or nearly uniform rightmost digits, whereas the falsified counts do not.[5]

Falsified counts (Notebook page 145)		Unquestioned counts supported by counter printouts (Notebook page 135)	
Numerator	Denominator	Numerator	Denominator
1078	251183	82267	170679
1770	217763	105584	190994
1091	225853	87592	181133
1434	238995	83341	197822
1247	241139	88426	172062
1131	260074	105068	194570
54350	220938	90707	150614

Table 1. Illustrative falsified and unquestioned counts from the respondent's laboratory notebook. Numerator (summation of reaction produced counts) and denominator (residual substrate count) are associated with a given clone, and activity is expressed by the ratio, numerator divided by denominator. Note that the 28 counts illustrated each contain from four to six digits.[4]

Chi-Square Results
For Falsified and Unquestioned Counts

Digits from 252 Falsified Counts

	Place 4	Place 3	Place 2	Place 1	Total
Number	185	250	252	252	939
Chi-Square	34.8	29.3	13.2	27.1	30.94
D. Freedom	9	9	9	9	9
Probability	.00006	.00058	.1521	.0013	.0003

Digits from 222 Unquestioned Counts

	Place 4	Place 3	Place 2	Place 1	Total
Number	195	218	222	222	857
Chi-Square	14.3	9.89	8.72	11.33	11.09
D. Freedom	9	9	9	9	9
Probability	.11	.36	.46	.25	.270

Table 2. Chi-square results for tests of uniformity of digit frequencies for falsified and unquestioned counts. The rightmost place is "Place 1"; the next place to left is "Place 2" etc. (Leftmost digits of numbers were excluded, so there are fewer "Place 4" digits than "Place 3," etc.)

Case 2: Unlikely Patterns in Rightmost Digits.

In this case, we again demonstrate the ability of uniformly distributed digits to distinguish questioned from unquestioned data. However, the digit analyses lead further to the identification of unlikely patterns in numbers that should not be related, given the purported experiment.

Table 3 (next page) reproduces the means and standard deviations from a published Table that was challenged by a coworker. Lipopolysaccharide extracts (LPS) were purified from endotoxin from various bacteria. The five rows in each half represent, respectively, different bacterial sources for the endotoxin. LPS was added at various concentrations to the cell cultures. Thus the five columns of the Table represent different levels of LPS (left to right, respectively: 5000, 500, 50, 5, and .5 ng/ml). The upper half of Table 3 represents cultures to which endotoxin and stimulator cells were added at the same time. The lower half represents cultures to which endotoxin was added 24 hours prior to the addition of stimulator cells. However, while supporting notebook data could be found for the first four columns, no supporting

Column 1		Column 2		Column 3		Column 4		Column 5	
Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
17697	1739	17399	1680	15085	1342	18262	2934	27191	1404
20164	3540	16746	1171	19397	1133	17889	3919	26999	7107
23323	3861	24154	722	19094	1340	28763	3373	28611	967
24474	4042	18918	4459	14224	828	24596	6327	29152	1407
29711	1519	21855	8458	23840	1695	29669	3222	28765	7104
24752	1455	22498	4591	21639	1347	32825	3063	70714	2106
32683	8535	26321	2753	20015	2020	34030	3917	68177	7155
43411	4682	41980	1705	34026	3906	47703	1894	66004	3924
26535	2349	41592	5699	31262	2796	54588	5065	74316	2192
33216	3762	37036	2071	27513	5062	32033	8307	71117	6817

Table 3. Published Table (Column 5 has questioned data).

	Column 1	Column 2	Column 3	Column 4	Column 5	Columns 1-4
Number	70	69	69	70	69	278
Chi-Square	8.57	5.93	8.54	9.14	26.22	4.45
D. Freedom	9	9	9	9	9	9
Probability	0.478	0.747	0.481	0.424	0.0019	0.880

Table 4. Tests of Uniformity of Digits for the Columns of the Published Table. Chi-Square tests of rightmost digits for, respectively, Columns 1 to 5 of the Published Table, and for Columns 1-4, together.

notebook data could be found for the questioned numbers in column 5.

Of statistical importance is the fact that means and standard deviations in this Table are reported to several places. Thus numbers are recorded with greater precision than the repeatability that the biological experiment allows, permitting a digit analysis.

The treatment of rightmost digits is the same as that for the previous case. Digits are analyzed in four places with no leftmost digit included in the analysis.

Only the digits of the questioned Column 5 are significantly different from uniform (p = .0019). Columns 1 to 4 separately are not different from uniform (the probability ranges from .424 to .747). In the aggregate, columns 1 to 4 are again not different from uniform (p = .88).

Based on the contrast between the digit distributions for the questioned Column 5 and the unquestioned columns, the complainant's assertion that the experiment for Column 5 was not done is strengthened.

Place 4	Place 3	Place 2	Place 1
1	4	0	4
7	1	0	7
	9	6	7
1	4	0	7
7	1	0	4

Table 5. Vertical Pattern of Digits

Furthermore, examination of the standard deviations in the upper half of Column 5 of Table 3 reveals a remarkable "vertical" pattern. These numbers should be statistically independent from row to row. However moving vertically downward at each digit place reveals a symmetrical recurrence of digits: 1,7, blank, 1, 7; 4, 1, 9, 4,1; then 0, 0, 6, 0, 0; and finally, 4, 7, 7, 7, 4 (Table 5).

The vertical pattern does not appear consistent with the five presumably statistically independent experiments depicted by the separate

Journal 1		Journal 2		Book	
Trauma patients		Cancer patients		Trauma patients	
26428	406	6428	406	116428	3406
7824	376	7824	376	17824	3761
24840	1107	24840	1107	124840	7107
26660	345	6501	355	116660	34511
7791	407	7906	348	17791	407
9276	1498	12016	1476	9276	1498

Table 6.

rows of Table 3. Such a pattern is consistent with the formation of the numbers after the outline of the published Table had been established.

Finally, to check for the possible existence of a pattern, three publications by the respondent (two journal articles and a book chapter) were examined. Examination of these publications reveals patterns of digits that are inconsistent with biological observations. Consider Table 6 (above), which contains numbers from tables in three different publications by the author, all for a similar experimental procedure.

In these three publications, rightmost digits that should not be reproducible are the same in the first and third rows, and they would be the same in the second row except for the arbitrary addition of a "1" after the "376" in the last column. Further, in the fifth row two of the standard deviations are "407" while the corresponding means are "7791" and "17791." Note that the standard deviation 7107 occurs in the book chapter and also in Column 5 of the published table already discussed. The respondent in this case agreed that the data were "flawed" and retracted the relevant articles.

Case 3: Banker's rounding and "odd" terminal digits

For the purposes of a genetic study, electro-physiological measurements of spontaneous "firings" (action potential spikes) of isolated muscle fibers were made. Initially, a firing was determined to occur whenever a peak on the recording of current equaled or exceeded 10 picoAmps. Since the spontaneous "firings" were infrequent, the continuous record of the electrical signal was not retained. Instead, an "event detection" device sent the time and the amplitude of the firing to Excel spreadsheets as a permanent record of the experiment.

To graph the activity of muscles from different genetic crosses, the firings of various amplitudes were accumulated into bins of 5-picoAmp width (10-15, 15-20, 20-25, etc), with accumulation continuing until some bin contained 100 firings.[6] The resulting frequency distribution represented the pattern of firings (for Experiment 1, see Figure 2, below, in which there are just over 100 events in the 20-25 bin).

Prior to publication, the respondent's coworkers thought that firings should only be defined as those peaks 20 picoAmps or greater. Thus they asked the respondent to prepare a new graph like that of Figure 2, but sampling only peaks 20 picoAmps or greater (i.e. resampling the Excel spreadsheet until some bin contained 100 such firings.)

The respondent submitted a new frequency graph that appeared like the first, but truncated at 20 rather than 10. Since one would expect the shape of the new graph (above 20 picoAmps) to differ, the coworkers questioned the result.

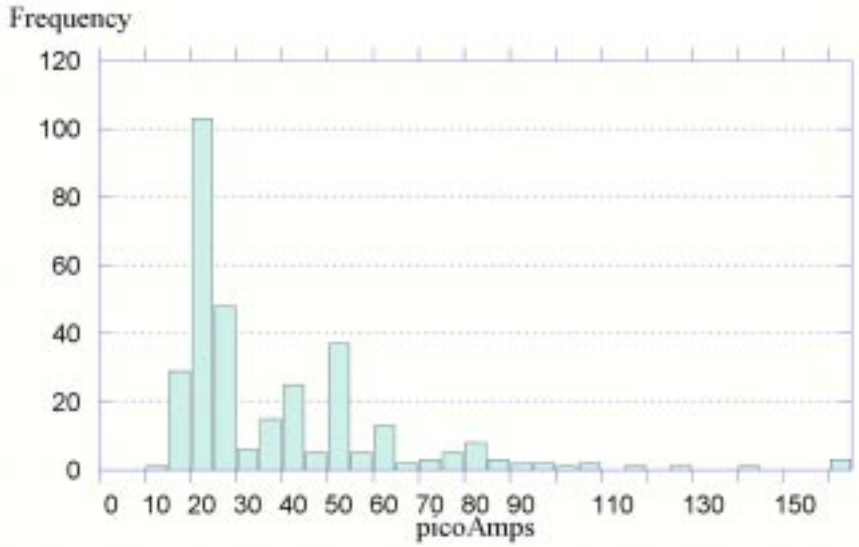

Figure 2. Binning of amplitudes into bins of 5-picoAmp width (initial 321 records of Experiment 1).

The respondent asserted that the new graph was not simply a truncated version of the first, but represented a fresh sampling of peaks greater than 20 picoAmps. He asserted that he had properly sampled the peaks in an Excel spreadsheet by counting beyond the initial 321 records on which the first graph (Figure 2) was based. The respondent furnished an Excel worksheet, "Experiment 1," of 551 records in support of the new graph. This worksheet contained the initial 321 records along with 230 additional records.

In addition to the Excel worksheet for Experiment 1, the respondent also provided a worksheet of unquestioned data "Experiment 2" with 1026 records. For Experiment 1 and the 10 picoAmp peaks, the initial 321 records of Experiment 1 are largely determined since the initial Figure 2 is known. Thus the last 230 records of Experiment 1 are more questionable. Since all 551 records were provided after the allegation, the opportunity existed to falsify or fabricate time points, but if falsifications occur, most would be expected in the last 230 records. Table 7, below, presents the first 12 records of Experiment 1.

It is interesting to note that all of the time values in Table 7 terminate in an even digit. The occurrence of only even time values can be explained by a long-used[7] practice sometimes known as "Banker's Rounding.[8]"

A simple explanation of the even terminal digits for time values is that two successive time-values are used in determining a peak, and the mid-point of the two is recorded. Thus when successive time values are added and divided by 2, the resulting terminal digit is 5 and would be rounded to an even digit, for example: (1000 + 1001)/2 = 1000.5 rounds to 1000, and (108.7 + 108.8)/2 = 108.75 round to 108.8. Therefore if numbers ending in 5 are rounded, only even numbers occur. The rounding of terminal 5's to the nearest even digit is the ANSI/IEEE standard[9] for rounding terminal 5's in computers. Examination of the terminal digits of the 1026 time values of the unquestioned data in Experiment 2 reveals no times ending in an odd digit. (The distribution of the 1026 penultimate digits of the times for Experiment 2 is not different from uniform (Chi-Square = 14.6, df = 9, p = .10).) In contrast, the questioned Experiment 1 contains time values that end in odd digits, reflecting insertions and alterations. In the initial 321 time points, six terminate in an odd digit (Figure 3). (The distribution of the 315 *penultimate* digits from the potentially unaltered *even* times is not different from uniform (Chi-Square = 8.14, df = 9, p = .52).)

Examination of the graph (Figure 4) of the final 230 records of Experiment 1 reveals many more (58) time values with odd terminal digits[10] than Figure 3. (The distribution of the 172 *penultimate* digits from the *even*, potentially unaltered, times is not different from uniform (Chi-Square = 12.3, df = 9, p = .20), whereas the distribution of 58 *penultimate* digits from falsified times ending in an odd digit deviates significantly from uniform (Chi-Square = 33.0, p = .00013).

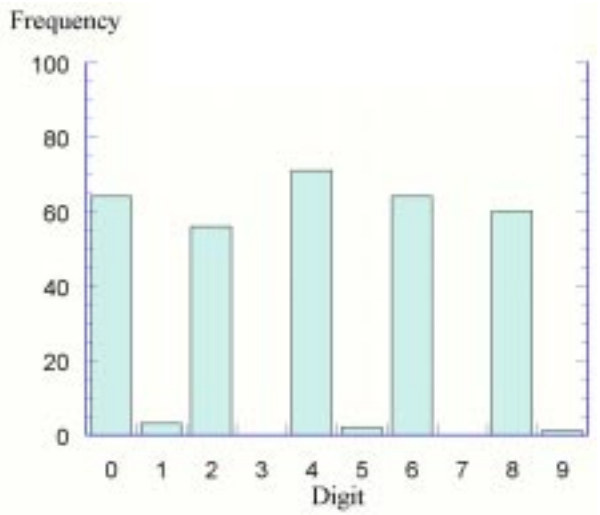

Figure 3. Experiment 1: first 321 time points; 321 terminal digits from 321 numbers. (Note presence of six odd digits.)

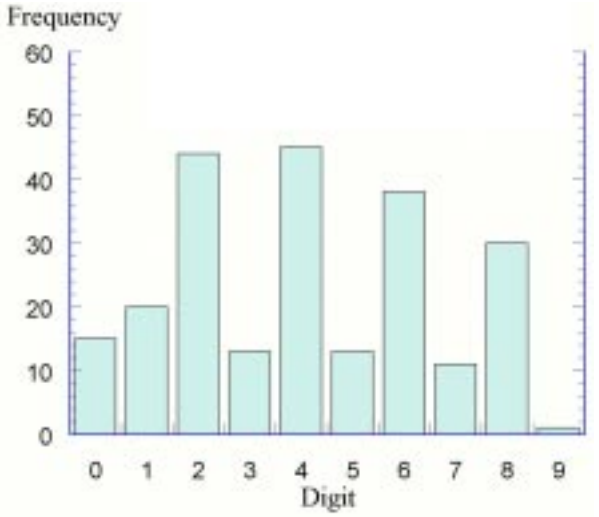

Figure 4. Experiment 1: last 230 time points; 230 terminal digits from 230 numbers. (Note presence of 58 odd digits.)

Experiment 1 - First 12 Records		
Time in Minutes	**Amplitude in picoAmps**	**Terminal Digit of Time**
0.0648	16.1	8
0.4904	22.7	4
0.4952	33.2	2
0.5398	19.8	8
0.9454	36.1	4
1.7182	44.4	2
2.6950	20.5	0
3.3626	19.3	6
3.7294	17.6	4
3.8586	14.9	6
4.3494	12.9	4
4.3712	45.4	2

Table 7. *The first 12 records of Experiment 1. Note that amplitudes include values less than 20, as expected. Also note that the terminal digit of the time is an "even" number for all 12 records.*

Many more time values terminate in odd digits in the final portion of Experiment 1, as expected if falsification occurred. The occurrence of time values ending in odd digits, mostly in the latter part of Experiment 1 (and the lack of uniformity of their penultimate digits) indicates data falsification. The timing of the occurrence of the minutes ending in odd digits is illustrated in Figures 5 and 6.

From Figure 6 it can be seen that not only do most of the odd time values occur in the last part of Experiment 1 (after minute 137.3006); it also appears from the denseness of the plot in the latter that the values immediately after this time point are quite close together. Further statistical tests of the intervals between events confirms the increased density in the latter part of Experiment 1, indicating the insertion of fabricated firing events.

Figure 5 (above). Experiment 2, unquestioned, 699 amplitudes (abs>20). (No amplitude is associated with an odd minutes.)

Figure 6 (right). Experiment 1, questioned; 371 amplitudes with abs>20, 52 with odd minutes. (Negative values, even minutes; positive values, odd minutes.)

	Rats					
	Weights-1	**Weights-2**	**Weights-3**	**Weights-4**	**Weights-5**	**Weights-6**
M-1	2.495	3.008	2.7515	4.631	2.250	3.4405
M-2	1.695	2.272	1.9835	3.019	0.702	1.8605
M-3	0.738	1.495	1.1165	1.768	0.843	1.3055
M-4	0.780	0.231	0.5055	0.394	0.085	0.2395
M-5	0.276	0.122	0.199	0.155	0.205	0.180
M-6	4.128	3.413	3.7705	2.261	1.187	1.724
M-7	1.131	1.224	1.1775	2.805	0.726	1.7655

Table 8. Portion of Excel spreadsheet with weights of muscles of rats 1-6. Note that some entries for columns Weights-3 and Weights-6 have four decimal digits and end in 5, whereas other entries have at most three decimal digits.

Case 4: One terminal digit too many

An investigator conducted studies on the effect of rhythmic contractions of skeletal muscle on blood flow using the hind limbs of rats. Blood flow was measured at rest and during nerve stimulation. In addition to measurements of blood flow, weights of individual skeletal muscles were recorded on data sheets. The experimental results for six rats were presented in a laboratory seminar. Sometime later a co-worker discovered that two of six data sheets were blank, and became suspicious that the measurements (blood flow/weights) had not been made for those rats. Suspicions were confirmed when frozen rat carcasses were checked. Although four had the hind limb muscles dissected, two were still intact and un-dissected. When confronted, the investigator (now respondent) admitted to falsifying data for two experimental animals. However, he subsequently withdrew the admission and denied the charges. The respondent stated that there was no evidence to support the claims that the research was falsified,[11] and that the university had not followed timely procedures.

The respondent presented to university officials blood flow and weight data for six rats on an Excel spreadsheet as well as purportedly original data sheets with handwritten entries for the muscle weights for six rats. Weights of 28 muscles and three other body parts for six rats extracted from the Excel printout are presented in Table 8.[12] Further weights as found in handwritten entries on separate data recording sheets for six rats are presented here in Table 9.

In Table 8, columns Weights-1, Weights-2, Weights-3 and Weights-6 correspond, respectively to columns 314-1, 314-2, 315-1 and 316-2 in Table 9. Thus the handwritten "original" data on the four data recording sheets (314-1, 314-2, 315-1 and 316-2) correspond to the columns labeled, respectively, Weights-1, Weights-2, Weights-3, and Weights-6 on the Excel spreadsheet. The columns Weights-4 and Weights-5 do not correspond to two additional data recording sheets labeled 315-2 and 316-1.

When values within a spreadsheet are calculated, rather than transcribed, the numbers may display more digits of accuracy than the original numbers that are the source of the calculated values. Therefore, looking for enhanced precision in spreadsheet numbers can indicate that certain numbers have been calculated or randomly generated by the spreadsheet software.

Since data are presented for six rats,

	Rats					
	314-1	**314-2**	**315-1**	**315-2**	**316-1**	**316-2**
M-1	2.495	3.008	2.725	3.859	3.479	3.440
M-2	1.695	2.272	1.984	2.087	1.881	1.861
M-3	0.738	1.495	1.117	1.464	1.320	1.306
M-4	0.780	0.231	0.506	0.269	0.242	0.240
M-5	0.276	0.122	0.199	0.202	0.182	0.180
M-6	4.128	3.413	3.771	1.933	1.743	1.724
M-7	1.131	1.224	1.178	1.980	1.785	1.766

Table 9. A portion of rat muscles weights from handwritten entries on six data recording sheets. Note that all numbers have a precision of three decimal places.

	Rat-3				Rat-6			
	Mean 1,2	**Weights-3**	**315-1**	**Difference**	**Mean 4,5**	**Weights-6**	**316-2**	**Difference**
M-1	2.7515	2.7515	2.725	0.0265	3.4405	3.4405	3.440	0.0005
M-2	1.9835	1.9835	1.984	-0.0005	1.8605	1.8605	1.861	-0.0005
M-3	1.1165	1.1165	1.117	-0.0005	1.3055	1.3055	1.306	-0.0005
M-4	0.5055	0.5055	0.506	-0.0005	0.2395	0.2395	0.240	-0.0005
M-5	0.199	0.199	0.199	0	0.18	0.18	0.180	0
M-6	3.7705	3.7705	3.771	-0.0005	1.724	1.724	1.724	0
M-7	1.1775	1.1775	1.178	-0.0005	1.7655	1.7655	1.766	-0.0005

Table 10. A Portion of the Weights for Rat 3 and Rat 6. The weights for Rat 3 are precisely the means of the respective weights for Rats 1 and 2. Additionally, the weights for Rat 3 correspond to three decimals to the handwritten weights for Rat 315-1. (The only exception is the weight for M-1 (shaded) where the rounded 2.752 is transcribed as 2.725. Correspondingly, the weights for Rat 6 are precisely the means of the respective weights for Rats 4 and 5. Additionally, the weights for Rat 6 correspond to three decimals to the handwritten weights for Rat 316-2, without exception.

and at most four allegedly were measured, the spreadsheet was evaluated for signs that some of the columns contained calculated values, rather than valid data entered from experimental records. The columns Weights-3 and Weights-6 in the Excel spreadsheet (Table 8) contain a number of entries that are recorded to one more decimal accuracy than the other columns (Weights-1, Weights-2, Weights-4, Weights-5). Additionally, these same entries for Weights-3 and Weights-6 contain one more digit than the purported original handwritten data as recorded on the sheets labeled 315-1 and 316-2 (Table 9). This extra precision could not occur from manual entry of the weights from the raw data sheets.

Instead, the presence of an extra digit indicates the possibility that these two columns represent calculated data. Further, where the extra digit occurs, it is always a "5." This indicates the calculation may have involved division by "2," suggesting that those numbers could be the means of two columns. (When the sum of the two numbers is even, there is no increase of the non-zero digits; however, when the sum is odd, division by 2 produces an additional "5" digit.)

In fact, the column Weights-3 is precisely the mean of columns Weights-1 and Weights-2 (see Table 10, below). Correspondingly, the column Weights-6 is the mean of columns Weights-4 and Weights-5 (Table 10).

Since these two columns are calculated on the spreadsheet, the "original" data on the recording sheets 315-1 and 316-2 are copied, respectively, from the spreadsheet-calculated columns Weights-3 and Weights-6. The only modification is that the "original" copied data are only transcribed to three-decimal accuracy as found on the (presumably) valid sheets labeled 314-1 and 314-2.

Lacking muscle-weight data for two rats, the respondent generated weights by twice forming means of measurements of other rats. The presence of the extra digit in the Excel spreadsheet provided the needed clue. When the respondent was shown that the two rats' weights were clearly produced as means, not measures, he accepted the finding of scientific misconduct.

Notes

1. 65 Federal Register 76260, December 6, 2000.
2. A theoretical discussion is found in J. E. Mosimann and M. V. Ratnaparkhi, "Uniform occurrence of digits for folded and mixture distributions on finite intervals," Communications in Statistics, 1996, **25***(2)*, pp *481-506*. Among other issues, this paper discusses approximations to continuous distributions by histogram-distributions for which the uniformity of terminal digits up to a specified place is known. Such theoretical issues are important, but our emphasis here is on direct comparison of questioned data with unquestioned data.
3. On May 1, 1995, the Maryland Lottery initiated a midday pick-3 drawing for weekdays. This is in addition to the nightly drawing. Thus there are more than 3,650 winning pick-3 numbers over the ten-year period. Maryland Lottery results may be found at the official website, *http://www.mdlottery.com*.
4. In all there are 474 counts: 252 admittedly falsified (notebook pages 141-152) and 222 unquestioned counts that are supported by counter printouts (notebook pages 104-106, 130-131, 134-135). Each count, falsified or unquestioned, contains from three to six digits. Digits were tested in four places, but no digit that was itself the leftmost digit was included in the analysis. Total analyses included 939 digits from 252 falsified numbers and 857 digits from 222 unquestioned numbers.

5. See "Data Fabrication: Can people generate Random Digits?" J. E. Mosimann, C. V. Wiseman and R. E. Edelman, *Accountability in Research*, **4**, *31-55*, 1995. This study shows that many people have difficulty fabricating random digits, even when trying to do so.
6. "Inverse" sampling until a certain number of a particular event occurs has a long history, particularly where rare events are to be studied. (*For example, see* J. E. Mosimann, "On the compound negative multinomial distribution and correlations among inversely sampled pollen counts," 1963, *Biometrika*, **50**, 47-54).
7. "It is conventional to round off to the nearest even digit when the number to be rounded is exactly half way between two successive digits." *pp*. 13-14, Paul S. Dwyer, *Linear Computations*, 1951, John Wiley & Sons Inc., *i-xi* , 1 – 344. (See also the next two footnotes.)
8. "PowerBASIC always rounds towards the closest even number. For example, both 1.5 and 2.5 would be rounded to 2. This is called banker's rounding. …" p. 169, *User's Guide*, 1997, PowerBASIC, Inc. 316 Mid Valley Center, Carmel, California, *i-vi*, 1-318.
9. ANSI/IEEE Std 854-1987, October 5, 1987, "ANSI" denotes the *American National Standards Institute* and "IEEE" denotes the *Institute of Electrical and Electronic Engineers, Inc.* "4.1 Round to Nearest. …if the two nearest representable values are equally near, the one with its least significant digit even shall be delivered." "5.4 Round Floating Point Number to Integral Value. …when rounding to nearest, if the difference between the unrounded operand and the rounded result is exactly one half, the rounded result is even."
10. 46 of these 58 time values that terminate in odd digits occur with amplitudes greater than 20 picoAmps. In the initial 321 records of Experiment 1, 6 of 6 odd time values occur with amplitudes greater than 20 picoAmps.
11. It is only after the respondent denied the charges and findings of the institution that the ORI demonstrated *which* two rats on the spreadsheet represented falsified data, and *the manner of falsification*.
12. The spreadsheet also contains columns of numbers representing blood pressure measurements and radioactive counts, some of which the university committee regarded as falsified. These are not presented here.